

Semi-Supervised Clustering Approach for P300 based BCI Speller Systems

Mandeep Kaur¹, A.K.Soni² and M. Qasim Rafiq³

¹School of Computer Science & Engg., Sharda University, Greater Noida, U.P., India, mandeephanzra@gmail.com

²School of Computer Science & Engg., Sharda University, Greater Noida, U.P., India, ak.soni@sharda.ac.in

³Dept. Computer Sc. & Engg. AMU, Aligarh, mqrafiq@hotmail.com

Abstract— The paper presents a k-means based semi-supervised clustering approach for recognizing and classifying P300 signals for BCI Speller System. P300 signals are proved to be the most suitable Event Related Potential (ERP) signal, used to develop the BCI systems. Due to non-stationary nature of ERP signals, the wavelet transform is the best analysis tool for extracting informative features from P300 signals. The focus of the research is on semi-supervised clustering as supervised clustering approach need large amount of labeled data for training, which is a tedious task. Hence works for small-labeled datasets to train classifiers. On the other hand, unsupervised clustering works when no prior information is available i.e. totally unlabeled data. Thus leads to low level of performance. The in-between solution is to use semi-supervised clustering, which uses a few labeled with large unlabeled data causes less trouble and time. The authors have selected and defined adhoc features and assumed the Clusters for small datasets. This motivates us to propose a novel approach that discovers the features embedded in P300 (EEG) signals, using an k-means based semi-supervised cluster classification using ensemble SVM.

Index Terms— P300 signal, Wavelet Transform, k-means, semi-supervised clustering, ensemble, support vector machines (SVMs).

I. INTRODUCTION

Brain computer interface (BCI) is a system that translates the electrophysiological activity of brain's nervous system into signals that can be interpreted by a device. Such systems are useful for able-bodied like game playing, multimedia applications etc and for unblessed or paralyzed patients like wheelchair, mind spelling, etc [1]. The electroencephalography (EEG) is a brain imaging technique for measuring the brain signals. The EEG based BCI systems are categorized into invasive and non-invasive. In invasive type of BCI system, microelectrodes are implanted into the skull of the user's brain. It gives high resolution signal, high signal-to-noise ratio but causes health problems. This reduces the use of invasive techniques during experimentation. There are various invasive techniques to record the brain's activity e.g. Electrocorticography (ECoG). On the hand, the non-invasive type of BCI system involves the recording of the brain electrical activity by means of placing the electrodes on the scalp. It gives low-resolution signal with high signal-to-noise ratio. However, these systems are easy to use, safe and cheaper. There are various non-invasive techniques like Electroencephalography (EEG). The most important advantage of EEG signal is that being a noninvasive technique it does not harm the subject [2]. The aim of research is to provide a system to assist unblessed in future and in addition putting an effort in appending the knowledge in the field of BCIs. Various P300 based

systems have been discussed in [3]. For this research, the mind spelling application has chosen that based on P300 signals. The P300 is a positive ERP signal with a latency of about 300ms after the presentation of rare or surprising, task-relevant stimuli. To evoke the P300, subjects are asked to observe two types of stimuli: one stimulus type (the oddball or target stimulus) that appears rarely and other stimulus type (the normal or non-target stimulus) that appears more often. A P300 signal can be detected in EEG as the target stimulus appears. Farwell and Donchin first proposed the paradigm for speller application, in 1988 [4]. Due to non-stationary nature of P300 (ERP) signals, wavelet transform is best signal analysis tool. The paper compared the accuracy result of various wavelet methods achieved for P300 based BCI systems [5]. Based on the comparison, in 2012 the Daubechies4 (db4) wavelet has achieved the highest accuracy of 97.50% than other techniques. The issues of using supervised clustering were highlighted in [5] and proposes a novel approach to employ k-means based semi-supervised learning for the classification by classifier ensemble SVM.

II. P300 SPELLER BCI SYSTEM PARADIGM

The objective of a BCI speller system is to design a system that enables a direct brain to character translation. P300 event-related brain potentials (ERPs) are very popular in BCI letter spelling applications. Farwell and Donchin designed one of the well-known P300 spellers in 1988. This section discusses P300 speller BCI system paradigm for which international datasets have collected and used. The aim of the speller system is to predict the correct character in each of the provided character selection epochs. The paradigm is a 6×6 matrix, shown in Figure.1, of alphanumeric characters [6-7].



Figure 1 User presented with a 6 by 6 matrix of characters. Every 175 ms (5.7 Hz) one of the rows, or one of the columns of the matrix was intensified [6-7] [<http://bbci.de/competition/iii/#datasets/descII>]

The international dataset used for testing the approach represents a complete set of P300 signals acquired by the group of Benjamin Blankertz at the Wadsworth Center with BCI2000 for the “BCI Competition III 2004”. The data acquired using BCI2000’s P3 speller paradigm, which is a flexible Brain-Computer Interface research and development platform. The user’s task is to focus attention on target character when corresponding row and column are intensified and count the number of flashing of the desired character silently. This was used only to keep subjects attention. The collected signals were band-pass filtered from 0.1-60Hz and digitized at 240Hz from two subjects in five sessions each. Each session consisted of a number of runs. In each run, the subject focused attention on a series of characters. Each row and column flashes randomly and successively for 100ms and at the rate of 5.7HZ. It means the time delay between every flash is 75ms with the frequency sampling of 240HZ. As we have 6 rows and 6 columns, it has to be 12 intensifications in every trial. 2 out of these 12 are target that are related to the desired character [8]. There were 15 trials for every character that ends to a total of 180 intensifications for a single character. After finishing 180 intensifications, there was a 2.5s time interval so that the user can prepare for the next character. The EEG recording was performed based on 10-20 system with sixty-four electrode channels in Figure 2. The P300 signals are retrieved from parietal lobe, various brain lobes are shown in Figure 3. Each row and column of the matrix was intensified for 100ms randomly. At any given moment, the user selects one of the letters or symbols that he wishes to communicate, and maintain a mental count of the number of times the row and the column of the chosen symbol are intensified. In response to this mental counting, a potential is elicited in the

brain. This procedure is called one trail, which is repeated 15 times for each selected letter or symbol. In each “trial” the subject is “communicating” a character by focusing attention on the cell containing the character [8].

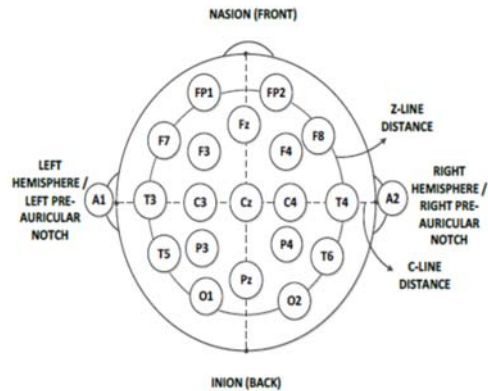


Figure2: 10-20 International System [9]

III. SEMI-SUPERVISED CLUSTERING

Clustering is a technique to partition a dataset into homogeneous clusters such that the data points in the same cluster are more similar to each other than in different clusters where classification is to label or classify a new unknown data from a collection of labeled, pre-classified, data. Clustering generally known as unsupervised learning where classification known as supervised learning. The term ‘learning’ states an algorithm that examines a set of points without examining any corresponding class/cluster label [11-12]. In various real and practical applications like bioinformatics, medical, pattern recognition etc, a large amount of unknown data is available than the labeled ones. To generate labeled data become a lengthy and slow process using unsupervised method, also is a tedious work to label all data using supervised method. Therefore, one may wish to use large dataset without labeling or generating data should employ semi-supervised learning. Semi-supervised learning is a technique of learning from a combination of labeled and unlabeled data. This can be used for both classification and clustering purpose. Semi-supervised classification uses labeled data along-with some unlabeled data to train the classifier where semi-supervised clustering, involves some labeled class data or pair wise constraints along with the unlabeled data to obtain better clustering. There are several semi-supervised classification algorithms like co-training, transductive support vector machines (SVMs), Expectation maximization etc for using unlabeled data to improve classification accuracy. The advantage of semi-supervised clustering is that the data categories (clusters) can generate from initial labeled data as well as extend and modify the existing ones to reflect other regularities in the data [10-15].

The clustering algorithms are classified into *generative* and *discriminative*. Generative is a parametric form of data generation is assumed and the goal in the maximum likelihood formulation is to find the parameters that maximize the probability (likelihood) of generation of the data given the model. Discriminative tries to cluster the data so as to maximize within-cluster similarity and minimize between-cluster similarity based on a particular similarity metric, where it is not necessary to consider an underlying parametric data generation model. Both can be implemented using Expectation maximization and k-means. In addition, k-means is a flat partitioning type of clustering algorithm that divides the data points into k partitions or clusters by grouping the similar features (usually Euclidean) and assigning each point to the cluster whose mean value on a set of x variables is nearest to it on that set. Furthermore, semi-supervised clustering algorithms categorize into similarity or distance-based or partially labeled data and search-based or pair constraint based methods. The former, used to classify the unlabeled data to the appropriate clusters using the known clusters where the later considers “Must-link constraints” require that two observations must be placed in the same cluster, and “cannot-link constraints” require that two observations must not be placed in the same cluster [13]. Another difference between the two is in distance-based method uses traditional clustering algorithm like k-means that uses a similarity metric where in search-based approaches, the clustering algorithm itself is modified so that user-provided labels or constraints are used to bias the search for an appropriate partitioning [10].

The distance-based Semi-supervised clustering is implemented using Seeded-K means (S-KMeans) and Constrained-K means (C-Kmeans). Both use labeled data to form initial clusters and constrain subsequent cluster assignment. The difference between two is in S-KMeans, the pre-defined labeling of the seed data may be changed during algorithm run where C-Kmeans is appropriate when the user want to change the label of seed data. In addition, C-Kmeans is suitable to use when the initial seed labeling is noise-free whereas S-KMeans is suitable with noisy seeds [11-12]. Due to the consideration of noise seeds in dataset, the proposed method employs generative, flat partitioning, distance-based Seed K-means semi-supervised clustering.

IV. PROPOSED FRAMEWORK

Data are collected using BCI 2000 competitions as discussed in section II. Data from only 10 channels have selected. A pre-processing module used to filter and normalize the data. Then, the wavelet features are extracted and projected to S-KMeans where the training of bagging based ensemble SVM performed.

In last, the pre-processed unknown data fed to the classifier for classification.

1. Load the training dataset: Size of the whole signal 85x7794x64. Two sets each A and B
2. Pre-processing
 - 2.1 Select the specific channel in which P300 signals are present: Out of 64 channel of each A and B, 10 channels are selected. Now size of A and B is 85x7794x10
 - 2.2 Lowpass and highpass butterworth filtering: low pass filtering at 30Hz and high pass filtering at 0.1Hz using an 8th order butterworth filter: size of A and B after filtering is 85x3897x10
 - 2.3 Coherence averaging: size of A and B after coherence averaging is 85x3897x10
 - 2.4 Independent Component Analysis includes De-correlation approach: estimate independent component one by one like in projection pursuit. Number of independent components to be estimated: equals the dimension of data step size: 1 Stopping criterion: 0.0001 Max number of iterations: 1000 Max number of iterations in fine tuning: 100 Percentage of samples used in one iteration: all Initial guess for corresponding mixing matrix: random After ICA, dimension of A and B is 85x3897x10
 - 2.5 Apply Principal Component Analysis, dimension of A and B is 3897x84x10
3. Apply wavelet filtering to extract the features to be trained using db4 wavelet. Now we have 2 outputs from PCA (A,B). Both are 2D. Apply 2D wavelet transform on each using 'db4' wavelets. We get 4 outputs: approx sig A, approx sig B, detail signal A, detail sig B, each of dimension 1952x45x10. Here, only approx. Coefficients are used as features.
4. Apply Seeded KMeans semi-supervised clustering on the obtained features. Consider random number of features from feature set to label the cluster. Assume that $k=2$ is known. On getting seed information, k-means will label the rest of the unlabeled features. [11] developed a generalization of k-means clustering for the problems where class labels are known. Let x_i and x_i' be observations from a data set with p features, and x_{ij} represents the value of the j th feature for observation i . Suppose further that there exists subsets S_1, S_2, \dots, S_K of the x_i 's such that $x_i \in S_k$ implies that observation i is known to belong to cluster k . (Here K denotes the number of clusters, which is also assumed to be known in this case.) Let $|S_k|$ denote the number of x_i 's in S_k . Also let $S = \bigcup_{k=1}^K S_k$. The algorithm proceeds as follows:
 - 4.1 For each feature j and cluster k , calculate the initial cluster means as follows:

$$\bar{x}_{kj} = \frac{1}{|S_k|} \sum_{x_i \in S_k} x_{ij}$$
 - 4.2 For each feature j and cluster k , calculate $|x_{kj}|$, the mean of feature j in cluster k .
 - 4.3 Repeat steps 2 and 3 until the algorithm converges [11-12].
5. The clusters are fed to Bagging based SVM Ensemble for training. Two types of kernel based SVM are used: polynomial and rbf.
6. The unknown feature vectors are then classified using trained Ensemble SVM.

V. IMPLEMENTATION

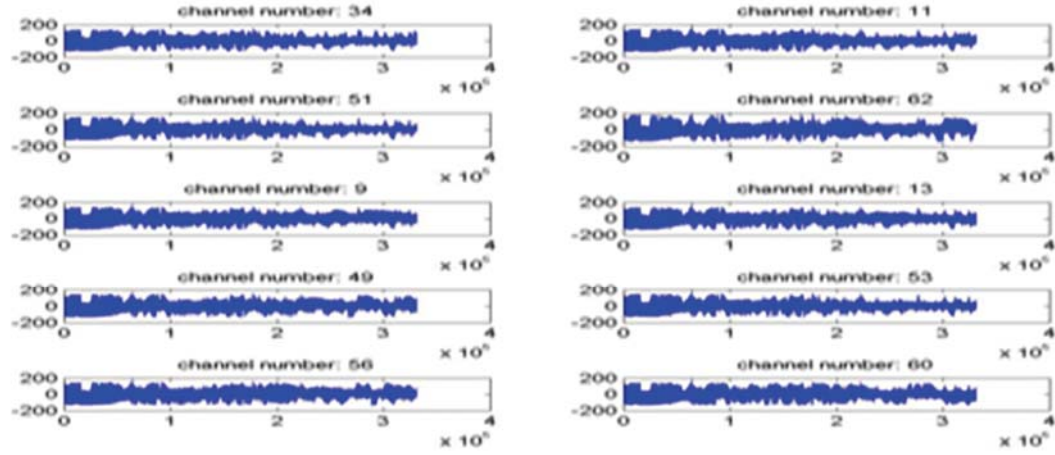
This section includes the results of 1, 2 and 3 steps (mentioned in section III) that includes loading of training data, pre-processing and feature extraction phase. BCI competition III data set II contained data collected

from two subjects A and B, recorded with the BCI2000 system, for P300 speller paradigm provided by Wadsworth Center, NYS Department of Health [13]. The data from each subject was comprised of 85 training symbols and 100 test symbols, with 15 trials per symbol, was sampled at 240 Hz. Here, the raw signals from only 10 selected channels are used. These channels Fz(34), Cz(11), Pz(51), Oz(62), P3(49), P4(53), PO7(56), PO8(60) / C3(9), C4(13) are shown in Figure 3. The results are considered for Subject A and B.

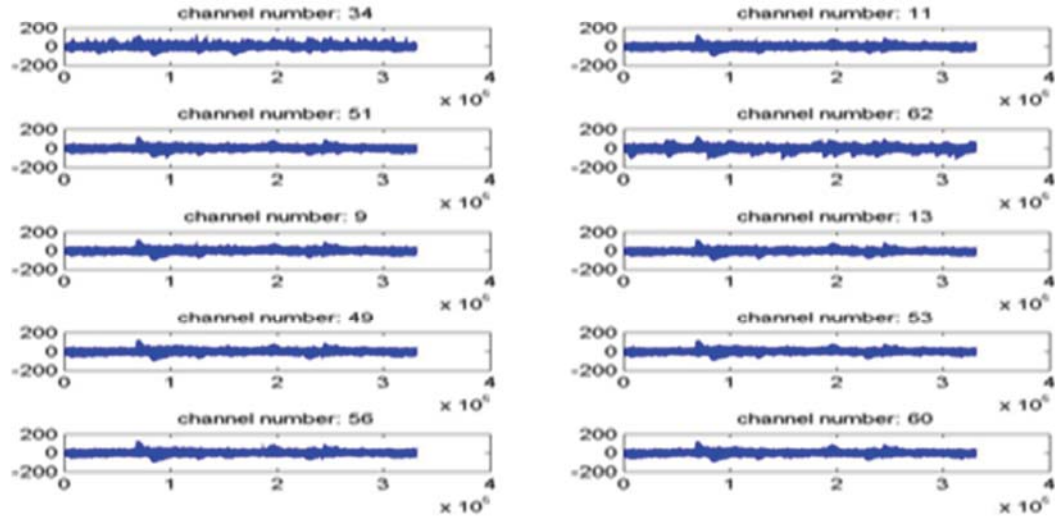
As the signals have very low amplitudes, they are highly affected by noise and artefacts. The bandwidth of 0.1-30.0 Hz covers the frequency range of important EEG rhythms (delta (0.5-4.0 Hz), theta (4.0-7.5 Hz), alpha (8.0-13.0 Hz), and beta (14.0-26.0 Hz)) as shown in Figure 4.

Then the other pre-processing methods Coherence Averaging, ICA, PCA and wavelet de-noising applied as shown in Figure 5, 6, 7 and 8 respectively.

For feature extraction phase, Daubechies (Db4) wavelet uses as the mother wavelet. The approximated coefficients used and concatenated into feature vectors.

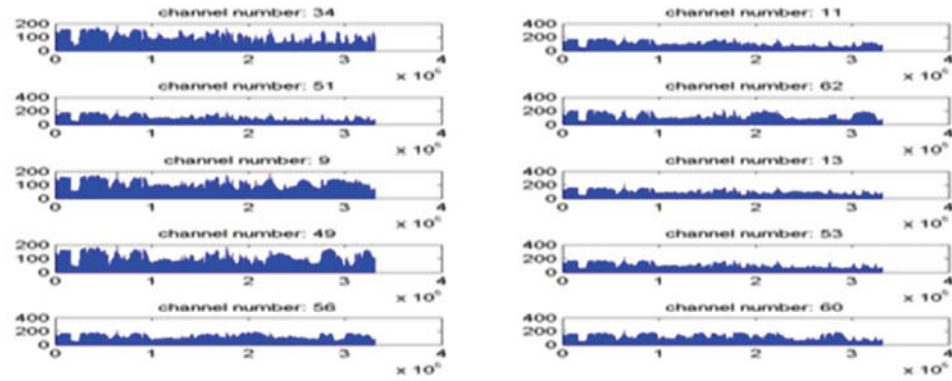


(a) Raw EEG signals (10 channels) from train dataset A

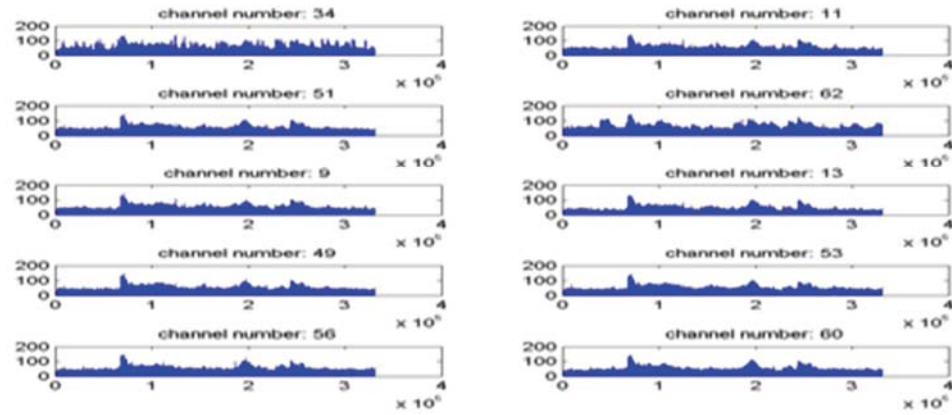


(b) Raw EEG signals (10 channels) from train dataset B

Figure 3: Raw EEG signals (10 channels) train dataset

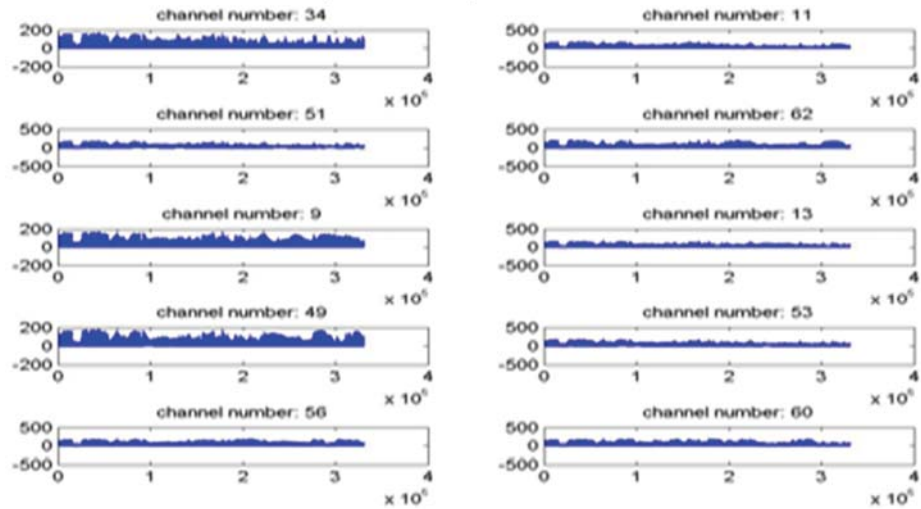


(a) Filtered Signals of train dataset A

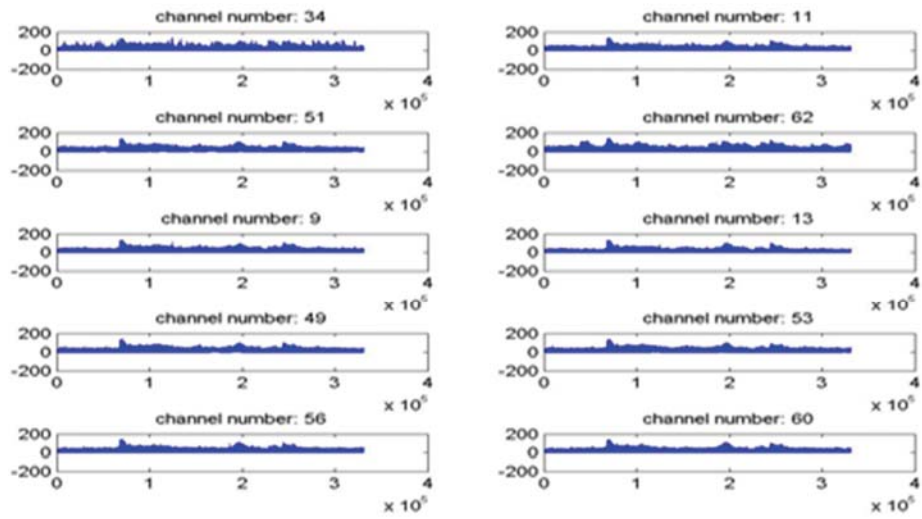


(b) Filtered Signals of train dataset B

Figure 4: Filtered signals of train dataset

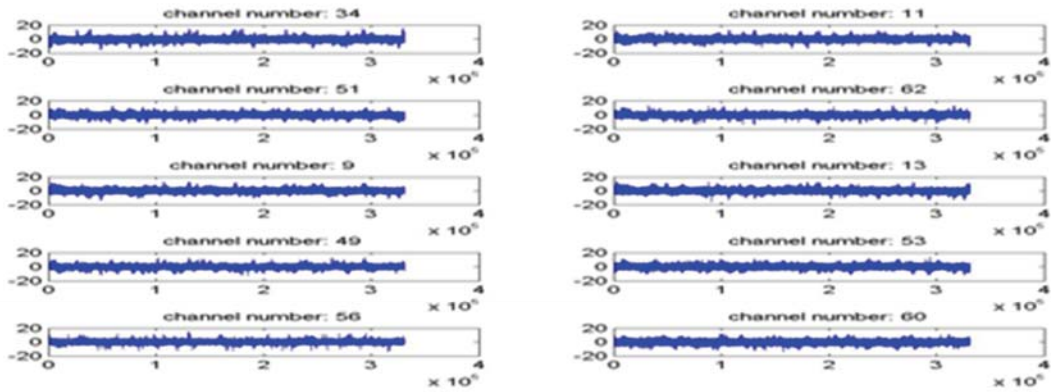


(a) Coherence Averaging for train dataset A

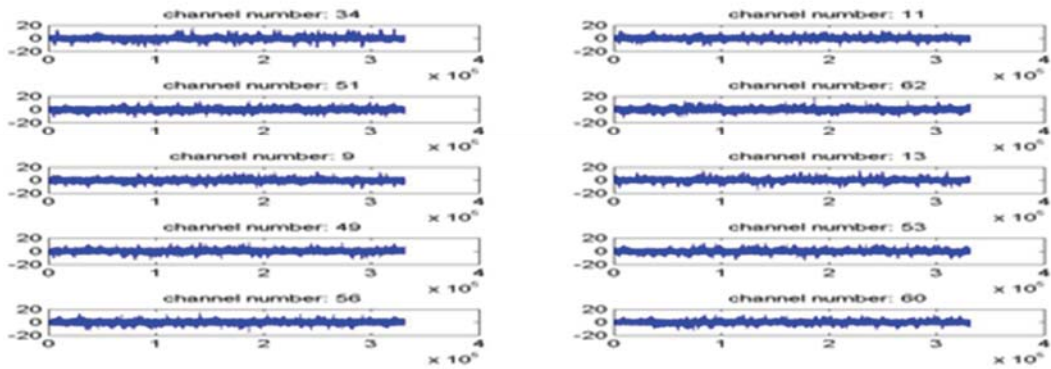


(b) Coherence Averaging for train dataset B

Figure 5: Signals after coherence averaging

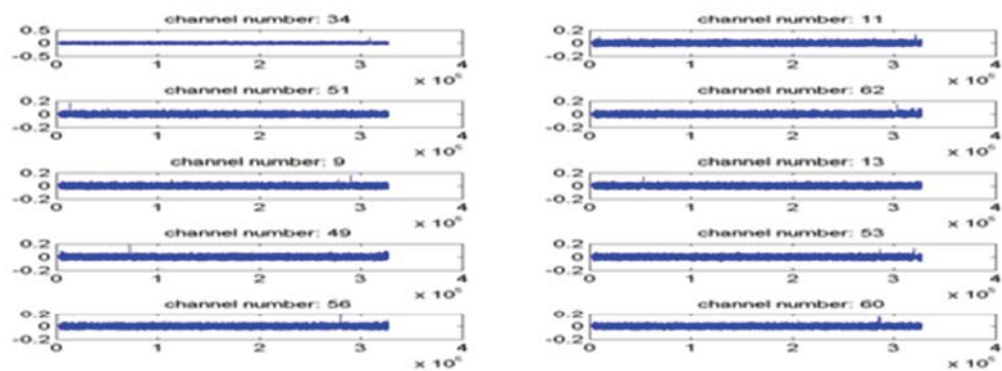


(a) ICA for train dataset A

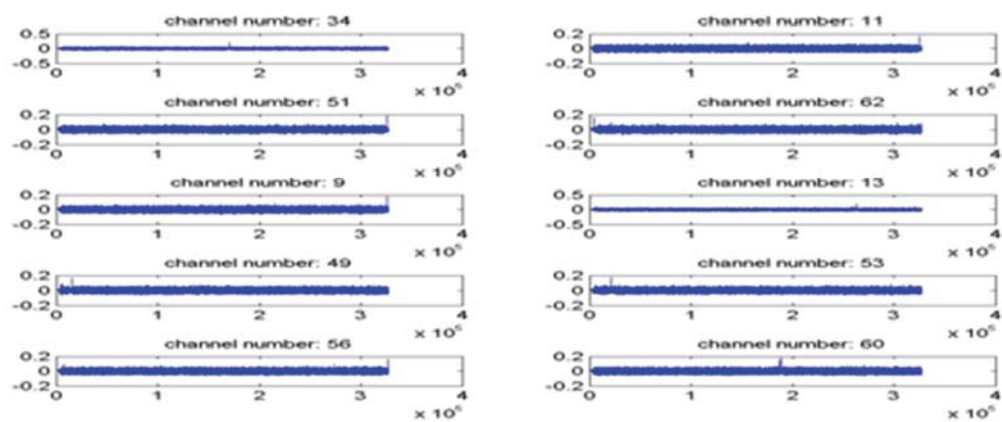


(b) ICA for train dataset B

Figure 6: Signals after ICA

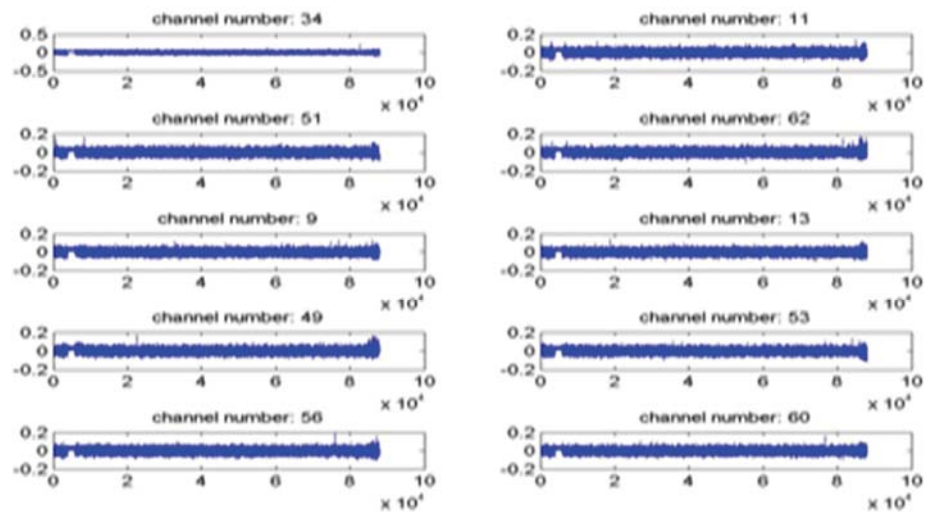


(a) PCA for train dataset A

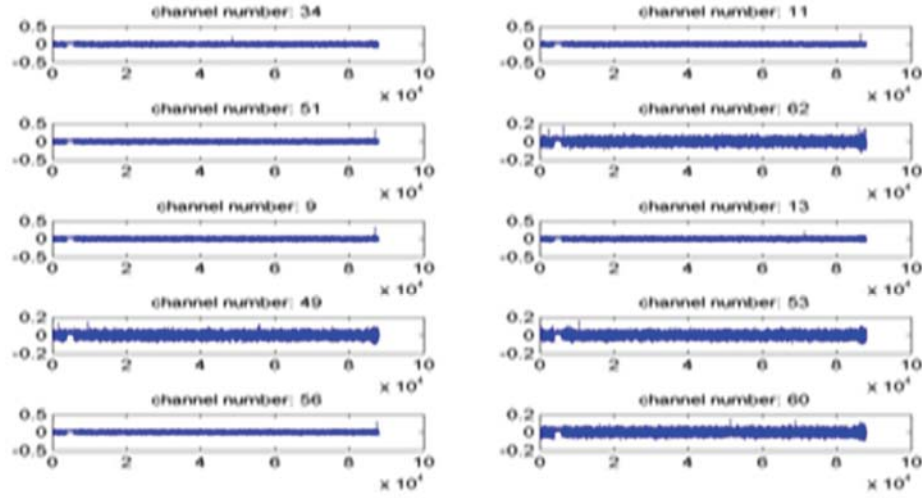


(b) PCA for train dataset B

Figure 7: Signals after PCA



(a) Approx coeff Db4 of train dataset A



(b) Approx coeff Db4 of train dataset B

Figure 8: Approximate wavelet coefficient of train dataset

VI. RESULT ANALYSIS

The analysis of P300 waveform in ongoing EEG performed using Signal-to-noise Ratio (SNR). The ERP signals are generally contaminated with other bio-signals EOG (Electrooculography), activity of the eyes, EMG (Electromyography) muscles, [14], and other interference like technical motion artefacts, the voltage to 60Hz, thermal noise, line noise and noise of the instrumentation itself [15]. Therefore, an ERP signal can be expressed as $x(t) = s(t) + n(t)$, where $x(t)$ is the contaminated signal, $s(t)$ is the required P300 signal and $n(t)$ is the noise or white noise with zero mean. Due to these, the amplitude of the ERP signal is much lower than the amplitude of the noise, this causes a significant decrease in signal-to-noise ratio (SNR) [16]. This causes obscurity in identifying P300 (ERP) signal in ongoing EEG signal. For this reason, it is desirable to decrease the noise in the signal to increase the signal to noise ratio to obtain target signal. Signal-to-noise ratio is defined as the power ratio between a signal wanted and noise. It is the ratio between the mean signal amplitude (evoked) and the standard error of the mean over trials. It can be derived from the formula

The signal-to-noise ratio (SNR) was defined as the ratio

$$\text{SNR} = P_{\text{signal}} / P_{\text{noise}} = \mu / \sigma$$

Where μ is the signal mean or expected value and σ is the standard deviation of the noise.

It can rewritten in decibels as

$$\text{SNR (dB)} = 10 \log_{10}(\sigma_s^2 / \sigma_n^2)$$

Where σ_s^2 is signal variance and σ_n^2 noise variance

// In Matlab

%Compute signal variance

var_sig = cov(sig);

%Calculate required noise variance

var_noise = var_sig / (10^(snr/10));

The aim of the pre-processing here is to enhance signal-to-noise ratio (SNR) and to eliminate artefact. The pre-processing involves coherence averaging of signals, ICA, PCA and Wavelet de-noising. The steps applied on BCI Competition III database II and motivating results appears as shown in Table 1 and Figure 9. This section represents the effect of averaging trials together on the training dataset. For example, the 5, 10, 15 and 20 target trials are averaged together with 5, 10, 15 and 20 non-target trials shown in Figure 10.

The SNR comparison of P300 windows from 250 to 350ms below 30 Hz for 5, 10, 15 and 20 averaged epochs (marked as “5, 10, 15, 20”) with the P300 component (in dB) for Subject A and B is shown in Figure 11.

TABLE I: COMPARING DIFFERENT METHODS ON ACHIEVED SIGNALS TO NOISE RATIOS OF 10 CHANNELS

Methods	SNR of Channel 1 Fz(34)		SNR of Channel 1 Cz(11)		SNR of Channel 1 Pz(51)		SNR of Channel 1 Oz(62)		SNR of Channel 1 P3(49)		SNR of Channel 1 P4(53)		SNR of Channel 1 PO7(56)		SNR of Channel 1 PO8(60)		SNR of Channel 1 C3(9)		SNR of Channel 1 C4(13)		Averages of SNR's	
Subject	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
Coherence Averaging Method	4.51	4.50	4.52	4.51	4.53	4.52	4.54	4.53	4.55	4.54	4.56	4.55	4.57	4.56	4.58	4.57	4.59	4.58	4.60	4.59	4.5339	4.5005
CA + ICA Method	4.56	4.55	4.57	4.56	4.58	4.57	4.59	4.58	4.60	4.59	4.61	4.60	4.62	4.61	4.63	4.62	4.64	4.63	4.65	4.64	4.5976	4.6108
CA + ICA + PCA	4.61	4.60	4.62	4.61	4.63	4.62	4.64	4.63	4.65	4.64	4.66	4.65	4.67	4.66	4.68	4.67	4.69	4.68	4.70	4.69	4.6867	4.6894
CA + ICA + PCA + Wavelet Denoising	4.66	4.65	4.67	4.66	4.68	4.67	4.69	4.68	4.70	4.69	4.71	4.70	4.72	4.71	4.73	4.72	4.74	4.73	4.75	4.74	4.7501	4.7339

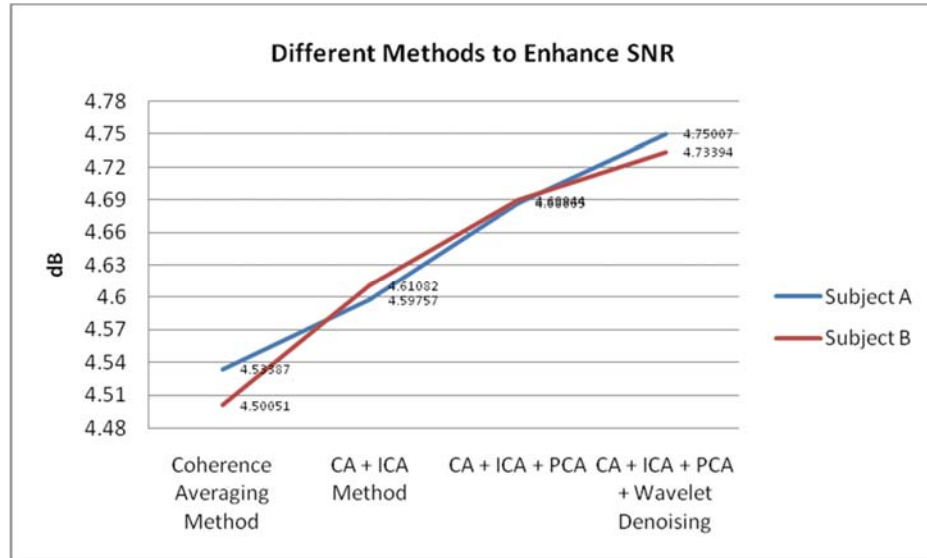


Figure 9: Comparing different methods on achieved Signals to Noise ratios of 10 channels

The remaining tests are under process and very soon will discuss with the final analysis of the proposed framework.

VII. CONCLUSIONS

Daubechies 4 (db4) is appropriate wavelet transform method for extracting the features of P300 signals. This paper use wavelet features for S-Kmeans based semi-supervised cluster classification using Ensemble SVM classifier. The goal is to train ensemble SVM from initial small-labeled training data and then extends for large datasets to predict the occurrence of P300 signal in ongoing EEG test dataset. Support vector machine is a very powerful machine learning algorithm as it does not have the over-fitting problem for the high dimensional data, can accommodate nonlinear classification by applying suitable kernels, computation is fast,

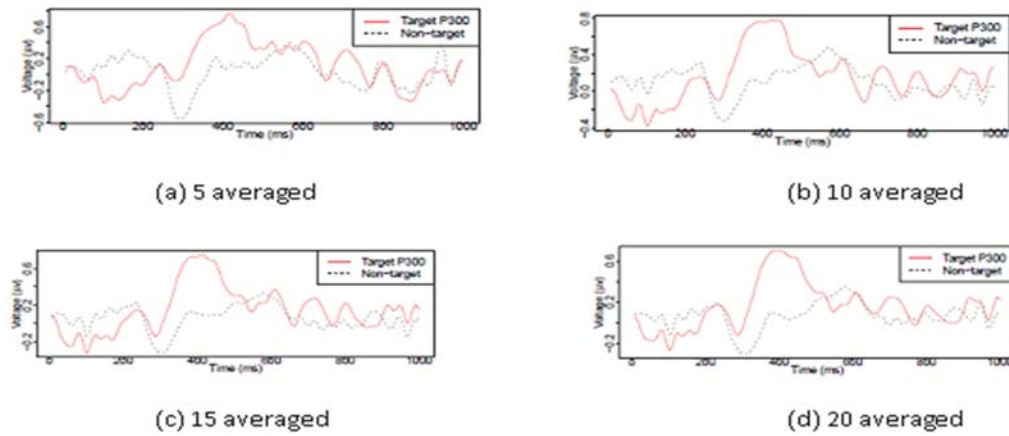


Figure 10: Averaging trials [17]

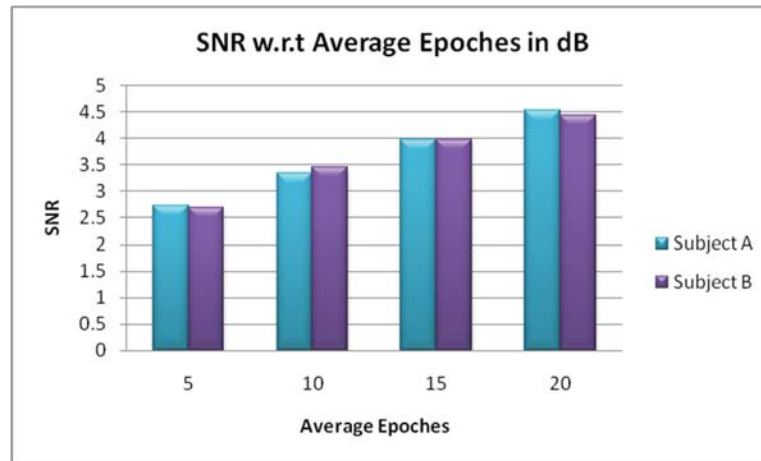


Figure 11: SNR w. r. t Averaging trials

extends to more than two classes, etc. The detailed description of the pre-processing phase has been discussed. The algorithm applied to BCI Competition 2004 dataset II and obtained the average SNR of 4.7501 dB for Subject A and 4.7339 dB Subject B. In addition, we reported the average SNR of 4.5338 dB with respect to average trial (20 trials). The aim of this research is to reduce the training time and improving classification accuracy, in the absence of enough labeled data, of P300 based speller systems. The future work emphasizes on further analysis of remaining implementation work.

REFERENCES

- [1] Nizar Ali Khemri, "P300 Wave Detection Using A Commercial Non-Invasive EEG Sensor: Reliability And Performance In Control Applications", Master of Sc. Thesis, Bachelor of Science in Computer Engineering, Al Fateh University, Tripoli, Libya 2006.
- [2] RezaFazel-Rezai, BrendanZ.Allison, ChristophGuger, EricW.Sellers, SonjaC.Kleih and AndreaKübler, "P300 brain computer interface:current challenges and emerging trends" Frontiers in Neuroengineering, vol. 5, Article 14, pp. 1-14, July 2012.
- [3] Mandeep Kaur, P. Ahmed, A.K. Soni, M. Qasim Rafiq, Wavelet Transform use for P300 Signal Clustering by Self-Organizing Map, International Journal of Scientific & Engineering Research, Volume 4, Issue 11, November-2013
- [4] LA Farwell and E. Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical neurophysiology*, 70(6):510–523, 1988.
- [5] Mandeep Kaur, P. Ahmed, M. Qasim Rafiq, Analysis of Extracting Distinct Functional Components of P300 using Wavelet Transform, Proceedings of 4th International Conference on Mathematical Models for Engineering Science

- (MMES '13) and the 2nd International Conference on Computers, Digital Communications and Computing (ICDCC '13), Brasov, Romania, pp. 57-62, June 1-3, 2013, ISBN: 978-1-61804-194-4.
- [6] H. Mirghasemi, R. Fazel-Rezai, M. B. Shamsollahi, "Analysis of P300 Classifiers in Brain Computer Interface Speller", Proceedings of the 28th IEEE EMBS Annual International Conference, New York City, USA, Aug 30-Sept 3, 2006
 - [7] Huiqi Li, Yuanqing Li, and Cuntai Guan, An Effective BCI Speller Based on Semi-supervised Learning, "Proceedings of the 28th IEEE EMBS Annual International Conference New York City, USA, Aug 30-Sept 3, 2006.
 - [8] <http://bbci.de/competition/iii/#datasets>, downloaded on 12 April, 2012.
 - [9] Mandeep Kaur, P Ahmed and Qasim M Rafiq. "Analyzing EEG based Neurological Phenomenon in BCI Systems", International Journal of Computer Applications, vol. 57, no. 17, pp.40-49, Nov 2012.
 - [10] Noam Shental, From Unsupervised to Semi-Supervised Learning: Algorithms and Applications, Thesis, THE HEBREW UNIVERSITY OF JERUSALEM November 2004
 - [11] Sugato Basu, Semi-supervised Clustering: Learning with Limited User Feedback, PhD Proposal, The University of Texas at Austin, Nov 2003 (Also Appears as Technical Report, UT-AI-TR-03-307)
 - [12] Sugato Basu, Semi-supervised Clustering: Probabilistic Models, Algorithms and Experiments, Thesis, The University of Texas at Austin, August 2005
 - [13] BCI 2004: <http://bbci.de/competition/iii/#datasets>
 - [14] Eric Bair, Semi-supervised clustering methods, Article first published online: 23 JUL 2013 DOI: 10.1002/wics.1270 © 2013 Wiley Periodicals, Inc.
 - [15] Osama Abu Abbas, Comparisons between Data Clustering Algorithms, International Arab Journal of Information Technology, vol. 5, no.3, July 2008.
 - [16] http://www.coli.uni-saarland.de/~crocker/Teaching/Connectionist/lecture13_4up.pdf , Frank Keller, Presentation on Clustering, Connectionist and Statistical Language Processing, Computerlinguistik Universit"at des Saarlandes.
 - [17] Cashero, Zachary, "Comparison of EEG preprocessing methods to improve the performance of the P300 speller", M.S. Thesis, Colorado State University, Fort Collins, Colorado, Sept 2011.